



*Working Paper n. 19 - 2010*

## **FREQUENCY OF EXAMINATIONS AND STUDENT ACHIEVEMENT IN A RANDOMIZED EXPERIMENT**

**Maria De Paola**

Dipartimento di Economia e Statistica  
Università della Calabria  
Ponte Pietro Bucci, Cubo 0/C  
Tel.: +39 0984 492459  
Fax: +39 0984 492421  
e-mail: m.depaola@unical.it

**Vincenzo Scoppa**

Dipartimento di Economia e Statistica  
Università della Calabria  
Ponte Pietro Bucci, Cubo 0/C  
Tel.: +39 0984 492464  
Fax: +39 0984 492421  
e-mail: v.scoppa@unical.it

---

**Novembre 2010**



# Frequency of Examinations and Student Achievement in a Randomized Experiment

*Maria De Paola, Vincenzo Scoppa\**

*Department of Economics and Statistics, University of Calabria*

*We carry out a randomized experiment involving undergraduate students enrolled at an Italian University attending an introductory economics class to evaluate the impact on achievement of examination frequency and interim feedback provision. Students in the treated group were allowed to undertake an intermediate exam and were informed about the results obtained, while students in the control group could only take the final exam. It emerges that students undertaking the intermediate exam perform better both in terms of probability of passing the exams and of grades obtained. High ability students appear to benefit more from the treatment. The experiment design allows us to disentangle “workload division or commitment” effects from “feedback provision” effects. We find that the estimated treatment impact is due exclusively to the first effect, while the feedback provision has no positive effect on performance. Finally, the better performance of treated students in targeted examinations seems not to be obtained at the expenses of results earned in other examinations.*

*Keywords: Education Production Function; Student Effort; Work Organization; Feedback Provision, Higher Education; Randomized Evaluation.*

*JEL classification: I21; J31; D82.*

## **1. Introduction**

A large and increasing body of economic literature has analyzed educational processes from a theoretical and an empirical point of view, trying to understand the role played by a number of factors, such as class-size, teacher abilities, peer group quality and parents' background in the determination of students outcomes.

Recently some works have tried to investigate the effects produced on student performance by teaching and evaluation practices. An important issue concerning this topic regards the optimal number of examinations, that is, whether it is better to test students more or less frequently, assigning them a smaller or a larger workload. On the one hand, when examinations are frequent and focused on a small number of issues students may find it easier to organize their work, with a positive effect on their learning process. In addition, students used to procrastinate their effort may end up studying more if there are more frequent deadlines.

Frequent examinations also offer students interim feedback of their results allowing them to

---

\* E-mail: [m.depaola@unical.it](mailto:m.depaola@unical.it); [v.scoppa@unical.it](mailto:v.scoppa@unical.it). We would like to thank the University of Calabria for providing access to the administrative data and the Dean of the Faculty of Economics for allowing us to carry out the experiment involving students enrolled at the Business and Administration Degree. We also thank Sara Laurita and Antonella Caparelli who assisted us with the use of the data.

know if their study effort has been appropriate and to become aware of their areas of strength and weakness. In fact, students may need a tangible way to measure their progress during a class.

On the other hand, when testing is too recurrent, students may not have enough time to deepen their knowledge and to understand the relationships among the range of concepts covered in a given subject. Moreover, they may be exposed to an excessive amount of stress.

The effects of test frequency and feedback provision on student achievement have been mainly investigated from a pedagogical and psychological point of view. Tuckman (1997), Deck (1998), Hattie and Timperley (2007), among others, show that more frequent testing produce positive effects on student performance. Ariely and Wertenbroch (2002) analyze the use of deadlines as a commitment device. According to their analysis, based on a sample of students enrolled at a MIT class, people have self-control problems and to avoid them choose costly deadlines that seem to improve their academic performance.

The economic literature has instead scarcely investigated this issue. Most of the existing literature examines feedback provision as an organizational design problem. In relative performance evaluations, for example, organizations decide whether to inform their employees about their relative standing at intermediate stages of the competition. In principle, information provision has ambiguous effects: performance might improve when workers who are obtaining a bad performance decide to work harder and try to avoid failure and when workers who are doing well become even more enthusiastic; on the other hand, performance tends to worsen when informed underdog become discouraged and workers getting a good performance, knowing that they are well ahead of the other colleagues, decide to shirk. These aspects have been analyzed from a theoretical point of view by a number of recent papers showing that feedback on past performance can affect current performance either directly – if past and current performances are substitutes or complements in the agent's utility function – or indirectly, by revealing information on the marginal return to current effort (Lizzeri *et al.*, 2002; Ederer, 2004; Perry, 2006; Yildirim, 2005; Aoyagi, 2010).

Empirical investigations trying to shed light on these effects are scant and lead to ambiguous results. Eriksson *et al.* (2008) show, through a laboratory experiment, that feedback on relative performance, regardless of the pay scheme used, does not improve performance. Instead, Freeman and Gelber (2010) find that individual performance improves when tournament participants are informed about their own and their competitors' past results.

Some works have tried to investigate these issues focusing on educational contexts. Bandiera *et al.* (2008) study the effect of providing university students with interim feedback information regarding their own performance. The authors show a positive and statistically significant effect of feedback provision on student final performance. In a similar vein, Azmat and Iriberry (2010), considering high school students in Spain, investigate the effect of informing students on whether they were performing above or below the class average. It emerges that students receiving this type of information obtain better grades.

In this paper we contribute to this emerging literature analyzing the effect of test frequency and feedback provision on a sample of Italian university students. Currently, in Italy University classes are typically organized in long modules offering about 60-80 hours of teaching activities. At the end of each class students take an exam and usually no intermediate assessment of student knowledge is undertaken. This organization is the result of the introduction of some new university rules, which have imposed a limit to the maximum number of exams needed to gain a First Level Degree. Before this change, teaching activity was organized in short modules (30-40 hours of teaching activity) and assessment was undertaken at the end of each module. The limit of a maximum number of exams introduced by the new law has forced Universities to reorganize their academic curricula and to unify short modules. The effects of this change have not been investigated yet. Nevertheless, if frequent examinations encourage students to study more and allow them to receive useful feedback then perhaps it would be worthwhile to put additional effort in trying to identify the optimal number of exams and to improve teaching organization.

To investigate the effects of test frequency and interim feedback on student performance we have conducted a randomized field experiment involving 344 undergraduate students enrolled at a middle sized Italian public University and attending an introductory economics class. Students participating at the experiment were randomly assigned to a control group and to a treatment group. Students in the treatment group were allowed to undertake an intermediate examination, while students in the control group were permitted to undertake exclusively the final examination at the end of the course program (as established by the University rules).

We firstly investigate “intention-to-treat” effects, considering as treated all the students assigned to the treatment group by the random procedure. It emerges that students in the treatment group perform significantly better than students in the control group in terms of probability of passing exams and of grades obtained. Subsequently, we investigate the impact of the effective participation to the treatment, using as an instrument the random assignment to the groups: adjusting the “intention-to-treat” effect for non-compliance leads to a stronger and highly significant effect of intermediate examination on student performance.

As the treatment effect can be seen as the combined result of a “feedback provision” effect and a “workload division or commitment” effect, we propose a framework to disentangle them. We find that the estimated effect is due exclusively to the “workload division or commitment” effect, while the feedback provision has no positive impact on performance.

Finally, we show that the improvement detected in the performance of treated students is not the result of a substitution effect. As a matter of fact, treated students did not obtain a worse performance in non targeted examinations compared to students in the control group.

The paper is organized as follows. In section 2 the design of the experiment is explained and some information on the Italian University system are provided. Section 3 shows the effects of examination frequency and feedback provision on the probability of passing exams and on grades

obtained by students. In Section 4, using an IV strategy, we investigate the treatment effects on students who have effectively undertaken the intermediate examination. In Section 5 we disentangle feedback effects from workload division or commitment effects. Section 6 investigates the existence of substitution effects. Section 7 concludes.

## ***2. Experiment Description and Data***

The experiment we conducted has involved 344 students enrolled at the Microeconomics and Macroeconomics classes offered by the First Level Degree Course in Business and Administration (BA hereafter) at the University of Calabria in the academic year 2009-2010. The University of Calabria is a middle-sized public university located in the South of Italy. It has currently about 34,000 students enrolled in different Degree Courses and at different levels of the Italian University system.

Since the 2001 reform, the Italian University system is organized around three main levels: First Level Degrees (3 years of legal duration), Second Level Degrees (2 years more) and Ph.D Degrees. In order to gain a First Level Degree students have to acquire a total of 180 credits. Students who have acquired a First Level Degree can undertake a Second Level Degree (acquiring 120 more credits). After having accomplished their Second Level Degree, students can enroll in a Ph.D. degree. The 2001 reform has introduced a credit system aimed to facilitate mutual recognition of degrees among European countries and has given autonomy to Universities. Among the unintended consequences of the reform there has been a proliferation of short classes. For example, at the BA Degree offered by the University of Calabria, a typical class allowed to acquire 5 credits, and students had to undertake more than 30 exams in order to acquire 180 credits. The introductory classes in Mathematics, Microeconomics, Macroeconomics, Statistics etc., were divided in two modules (5 credits each) assessed through two different exams at the end of each teaching period. In order to limit the proliferation of classes, a new rule, known as “DM 270”, coming into effect in the academic year 2008/09, has imposed a maximum of 20 exams for First Level Degrees leaving unchanged the number of credits to acquire (180).

At the University of Calabria the “DM 270” has led to a reorganization of the academic curriculum in order to reduce the number of exams students had to undertake. In particular, some classes, such as Macroeconomics and Microeconomics, initially split in two modules were unified in a single class (10 credits) and students were required to undertake a unique exam.

One of the ideas behind the “DM 270” was to give students more time to prepare for exams and avoid them the stress deriving from sitting for a large number of examinations. However, as explained above, reducing the number of exams might also produce some negative effects on student performance.

By comparing the average performance of students enrolled at the BA Degree, immediately before and after the introduction of “DM 270”, it emerges a drop in the number of credits acquired by

students. Students enrolled in 2007/2008, under the old regime, have acquired 38.9 credits during their first year, while their counterparts enrolled in 2008/2009, under the new system, have acquired only 32.4 credits. However, these figures might be driven by temporal trends and unobserved changes in student and instructors' characteristics. Then, in order to try to shed light on this issue, we have decided to undertake the experiment described in this paper.

At the beginning of the Microeconomics and Macroeconomics classes (in March 2010) students were informed of the experiment both through presentations during the teaching hours and through a letter, sent to all students, explaining the format of the experiment.

We asked students to register for joining the experiment. We did not consider in the experiment the small fraction (around 10%) of students who did not register.<sup>1</sup> On the basis of the available administrative information on students' characteristics, we proceeded to the stratification of students participating to the experiment according to the following variables: class attended (Microeconomics or Macroeconomics); gender; type of High School attended (3 categories: Lyceum; Technical schools; Vocational and other types of schools); final grade obtained at High School (split in 4 categories corresponding to quartiles).

Following this procedure the 344 students were allocated to 48 non null groups. Within each group, one half of students was randomly assigned to the treatment group – allowed to take the intermediate exam – and the other half was assigned to the control group, which could take a unique exam at the end of the class, without an intermediate exam.<sup>2</sup> We ended up with 172 students assigned to the treatment group, and 172 to the control group.

The random assignment procedure was carried out at the presence of students. They were also informed by e-mail of their assignment status and the list of students belonging to the treatment and control group was published on the classes' web-pages.

The Microeconomics and Macroeconomics classes were taught to students enrolled, respectively, at the first and second year of the BA Degree. Both these classes started in March 2010 and lasted until June. Each class program consisted in 60 hours of teaching activity and 20 hours of laboratory. Treatment and control groups attended the class in the same room, at the same time and with the same instructor and teaching material.

After the first 30 hours of teaching activity (and 10 hours of laboratory) there was a break of two weeks (May 2010). The classes teaching programs were then naturally divided in two parts (that is, Microeconomics-1 and Microeconomics-2 and Macroeconomics-1 and Macroeconomics-2). During the break period, students in the treatment group were allowed to undertake an intermediate exam covering respectively the Microeconomics-1 and the Macroeconomics-1 program. The intermediate exam consisted in 30 multiple choice questions and lasted one hour. 77% of students in the treatment

---

<sup>1</sup> These students may be low motivated students, not attending classes and not particularly involved with studying activities.

<sup>2</sup> When the number of students included in each stratified group was not even, one student was first assigned randomly to the treatment or control group.

group have effectively undertaken the test.

Teaching activities re-started in the mid of May and lasted for other 30 hours (+10 hours of laboratory), until the end of June. In July students in the treatment group were required to complete their examination by taking an exam covering exclusively the second part of the teaching program (respectively, Microeconomics-2 and Macroeconomics-2), consisting in 30 multiple choice questions (in one hour). The final grade obtained by these students was given by the average of the grades earned at the first and at the second part of the exam.

Students in the control group have, instead, undertaken the exam covering the whole course program (the first plus the second part). This exam was held in July and has consisted in 30+30 multiple choice questions. These students had the double of time allowed to students in the treatment groups to complete the entire examination (two hours broken by an interval of 30 minutes).

As regards the first module, since each exam question was randomly selected from a large test-bank, the exam for treated students can be considered equivalent to the exam for students in the control group. The second module was identical for treated and control students.

We had a number of non-complier students: 22 treated students never shown (neither to the intermediate nor to the whole examination); 43 treated students shifted to the control group (17 did not undertake the intermediate exam, while 26 chose to undertake the entire exam after having participated at the intermediate exam).<sup>3</sup> Finally, 27 students took the intermediate exam, but did not complete the exam undertaking the second part.

In this first part of our analysis we focus on “intention-to-treat” effects, considering as treated all the 172 students randomly selected. In sections 4 and 5 we proceed both by using an instrumental variable strategy and by defining a more restricted sample of students.

We measure student performance considering both the probability of passing the target examinations and the grades obtained. In the Italian system, passing grades range from 18 to “30 cum laude”, which we consider equal to 31. We observe both grades in passed examinations (18-31) and in failed examinations. Only exams undertaken until the 31<sup>st</sup> of July 2010 were taken into account in determining student performance.

Table 1 provides descriptive statistics for our sample of students. About 61% of students were female. 44% of students attended the Macroeconomics class while 56% attended Microeconomics. High School Grade ranged from 60 (the minimum passing grade) to 100 (the maximum grade), with a mean of 88.8. Students mainly came from Technical Schools (45%) and Lyceums (about 42%). As an additional measure of student ability we have considered the number of credits acquired by students until the beginning of the experiment, *Credits*, and to make comparable the two cohorts of students involved in the experiment we have divided the number of credits by the number of exam sessions they had available (1 for first year students and 3 for second year students). The average number of

---

<sup>3</sup> We were forced by the University rules to allow students willing to undertake the entire exam after having participated at the intermediate exam to do so (these are students who obtained a very low grade at the intermediate exam).

credits acquired in a semester was 14.24. About 55% of students came from the same province (NUTS-3) where the University is located. 14% of sample students did not enroll at University immediately after High School graduation, but a year or more later (*Late Enrollment*).

At the end of exam session, 35% of students passed the exams considered in this experiment. The average grade (ranging from 1 to 31 is 15.4). The average grade for students passing the exams was 22.6.

**Table 1. Descriptive Statistics**

Variables	Obs	Mean	Std. Dev.	Min	Max
Pass	344	0.348	0.477	0	1
Grade	236	15.371	8.571	1	31
Treatment (Intermediate Exam)	344	0.500	0.501	0	1
Macroeconomics	344	0.439	0.497	0	1
Female	344	0.613	0.488	0	1
Credits	344	14.245	6.865	0	28.667
High School Grade	344	88.828	8.479	68	100
Technical Schools	344	0.453	0.499	0	1
Lyceum	344	0.416	0.494	0	1
Late Enrollment	344	0.140	0.347	0	1
Resident near University	344	0.558	0.497	0	1

Notes: Grades in each class ranges from 18 to “30 cum laude” (set equal to 31). High School Grade ranges from 60 to 100.

In the first two columns of Table 2 are reported, by treatment groups, means for a number of individual characteristics. In the third column we report differences in means between treatment and control groups (standard errors are reported in parentheses). Results show that the randomization has been successful in creating comparable treatment and control groups along the observable characteristics: there are no significant differences by treatment status in class, gender, number of credits acquired, High School Grade, type of High School attended, Late Enrollment and place of residence.

**Table 2. Student characteristics across treatment and control groups**

	Means		Differences (s.e.)
	Treatment	Control	Treatment v. Control
Macroeconomics	0.442	0.436	0.006 (0.054)
Female	0.610	0.616	-0.006 (0.053)
Credits	14.031	14.459	-0.428 (0.741)
High School Grade	88.564	89.093	-0.529 (0.915)
Technical Schools	0.453	0.453	0.000 (0.054)
Lyceum	0.419	0.413	0.006 (0.053)
Late Enrollment	0.157	0.122	0.035 (0.037)
Resident near University	0.552	0.564	-0.012 (0.054)

Notes: Standard errors are reported in parentheses.



### 3. Treatment Effects on Student Achievement

In this Section we analyze the effects of treatment both on the student probability of passing examinations and on grades obtained. We focus on “intention-to-treat” effects considering as treated all students randomly assigned to the treated group.

#### 3.1. The Effect of Treatment on the Probability of Passing the Exam

We use a probit model to estimate the probability of passing the examination targeted in the experiment (either Microeconomics or Macroeconomics):

$$[1] \quad \Pr(Pass_i = 1 | Treatment_i, X_i, \mu_i) = \Phi(\beta_0 + \beta_1(Treatment_i) + \phi X_i + \mu_i)$$

Our dependent variable  $Pass_i$  takes the value of one if student  $i$  passed the exam, and 0 otherwise,  $Treatment_i$  is a dummy variable which takes value of 1 if student  $i$  was assigned to the treatment group while takes value of 0 if  $i$  belong to the control group,  $X_i$  is a vector of the individual characteristics of  $i$  (measures of his/her ability and personal characteristics),  $\mu_i$  is a dummy for the Macroeconomics class,  $\varepsilon_i$  is an error term capturing idiosyncratic shocks or unobserved student characteristics.

For each student we observe separately the outcome of the two modules composing each exam. However, in this section we have organized data at student level taking as a measure of student performance the average outcome at the two modules. This implies that we consider the exam as passed when the average grade obtained by the student is equal or higher than 18.

The coefficient on *Treatment* captures the “intention-to treat” effects on student performance (with respect to the control group). Estimates are reported in Table 3. In all the specifications, standard errors are robust to heteroskedasticity.

In column 1 we regress  $Pass_i$  on the *Treatment* and *Macroeconomics* dummies, without other controls. It emerges that students in the treatment group have a probability of about 15 percentage points higher to pass the exam. The coefficient is significant at the 1 percent level. In column 2 we control for individual characteristics: the probability of passing the exam increases by 20 percentage points for students in the treatment group.

In columns 3 and 4 we run separate regressions for the Macroeconomics and Microeconomics classes. The effect appears relevant for both classes. In columns 5 and 6, showing separately results for males and females, it seems that males react more than females to the treatment, but when we use an interaction term to test the statistical significance of this difference we are not able to reject the null hypothesis (results not reported).

As robustness checks we have also used a linear probability model. OLS estimates are very

similar to the probit estimates presented in Table 3 and are not reported.

**Table 3. Estimates of the treatment effect on the probability of passing the exam. Dependent variable: Pass**

	(1) All	(2) All	(3) Macro	(4) Micro	(5) Females	(6) Males
Treatment (Intermediate Exam)	0.153*** (0.051)	0.205*** (0.054)	0.290*** (0.077)	0.136* (0.074)	0.161** (0.072)	0.262*** (0.083)
Macroeconomics	-0.071 (0.051)	-0.187*** (0.057)			-0.198** (0.077)	-0.176** (0.085)
Female		-0.042 (0.057)	-0.041 (0.086)	-0.041 (0.078)		
High School Grade		0.013*** (0.004)	0.011** (0.005)	0.015*** (0.005)	0.012*** (0.004)	0.012** (0.006)
Credits		0.033*** (0.005)	0.028*** (0.009)	0.036*** (0.006)	0.040*** (0.007)	0.028*** (0.008)
Technical Schools		0.005 (0.091)	0.029 (0.130)	-0.002 (0.130)	0.210* (0.127)	-0.241* (0.131)
Lyceum		0.112 (0.093)	0.162 (0.132)	0.089 (0.133)	0.270** (0.133)	-0.127 (0.129)
Late Enrollment		0.021 (0.088)	0.089 (0.164)	-0.013 (0.103)	0.018 (0.131)	0.056 (0.131)
Resident near University		0.058 (0.056)	0.053 (0.081)	0.068 (0.075)	-0.057 (0.070)	0.231*** (0.089)
Observations	344	344	151	193	211	133
Pseudo R-squared	0.024	0.244	0.256	0.247	0.347	0.168
Log-likelihood	-217.180	-168.159	-69.703	-96.345	-88.383	-72.322

Notes: The Table reports marginal effects from probit estimates evaluated at the mean values of the explanatory variables in the sample. Robust standard errors are reported in parentheses. The symbols \*\*\*, \*\*, \* indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

The effects of control variables are consistent with the findings emerging from the education literature. Both our measures of student predetermined ability, *High School Grade* and *Credits*, produce a positive and statistically significant effect on academic performance. Once we control for these measures of abilities neither the dummy *Lyceum* nor the dummy *Technical Schools* produce a significant effect. The dummy *Female* is not statistically significant. The probability of passing the Macroeconomics examination turns out to be lower than the probability of passing Microeconomics.

### 3.2. The Effect of Treatment on Grades

In this section we estimate the effect of the treatment on student grades. Our dependent variable is the grade obtained by student  $i$  at the exam. To take into account the fact that *Grade* is censored, since a number of students did not sit for the exam and we do not have information on their performance, we use a Tobit model expressing the observed outcome  $Grade_i$  in terms of a latent variable  $Grade_i^*$ :

$$[2] \quad Grade_i^* = \beta_0 + \beta_1(Treatment_i) + \phi X_i + \mu_i + \varepsilon_i$$

$$Grade_i = \max(0, Grade_i^*)$$

We set equal to zero the variable *Grade* for absent students.

Table 4 reports Tobit estimates using the same specifications as in Table 3 (the lower limit has been set equal to 0 since the minimum grade observed is 1). In Panel A are reported the marginal effects of explanatory variables on the expected grade conditional on being uncensored, that is, given that the student has sit for the exam. In all the specifications it emerges that taking an intermediate exam has a positive and statistically significant impact on student achievement: the expected grade increases by about 5 points when controlling for individual characteristics.

On the basis of the Tobit estimations, we have also determined the marginal effect of treatment on the probability that a student will undertake the exam (that is, the effect on the probability that the grade is greater than 0). These effects are reported in Panel B of Table 4. Undertaking the intermediate examination increases the probability of sitting for the exam by 25 percentage points.

Treatment effects do not differ much according to the type of class attended (columns 3 and 4) by the students and according to gender (columns 5 and 6).<sup>4</sup>

---

<sup>4</sup> We have also estimated a Tobit model with left and right censoring using 31 as the upper limit for right censoring. Results (not reported) are very similar.

**Table 4. Tobit estimates of the Treatment Effect on Grades**

<b>Panel A. Marginal Effects: Conditional on Being Uncensored</b>							
	(1)	(2)	(3)	(4)	(5)	(6)	
	All	All	Macro	Micro	Females	Males	
Treatment (Intermediate Exam)	4.489***	4.993***	5.124***	4.839***	4.890***	4.901***	
	(0.782)	(0.702)	(1.023)	(0.966)	(0.856)	(1.211)	
Macroeconomics	-1.057	-2.577***			-2.920***	-1.875	
	(0.777)	(0.712)			(0.874)	(1.229)	
Female		-0.215	-0.593	0.259			
		(0.709)	(1.075)	(0.957)			
High School Grade		0.088*	0.019	0.145**	0.045	0.165**	
		(0.045)	(0.071)	(0.061)	(0.055)	(0.081)	
Credits		0.528***	0.510***	0.542***	0.593***	0.466***	
		(0.057)	(0.091)	(0.074)	(0.068)	(0.109)	
Technical Schools		-0.808	-0.344	-1.247	0.439	-3.083	
		(1.115)	(1.563)	(1.585)	(1.396)	(1.913)	
Lyceum		1.075	1.565	0.697	1.835	-0.516	
		(1.114)	(1.558)	(1.585)	(1.403)	(1.884)	
Late Enrollment		0.863	0.711	0.833	1.046	0.742	
		(0.987)	(1.536)	(1.315)	(1.247)	(1.616)	
Resident near University		0.144	-0.008	0.195	-0.948	1.859	
		(0.708)	(1.041)	(0.975)	(0.865)	(1.285)	
<b>Panel B. Marginal Effects: Probability Uncensored</b>							
Treatment (Intermediate Exam)	0.224***	0.257***	0.296***	0.226***	0.253***	0.250***	
	(0.040)	(0.037)	(0.061)	(0.046)	(0.045)	(0.063)	
Macroeconomics	-0.054	-0.139***			-0.157***	-0.100	
	(0.039)	(0.037)			(0.046)	(0.064)	
Female		-0.011	-0.034	0.012			
		(0.037)	(0.064)	(0.045)			
High School Grade		0.005*	0.001	0.007**	0.002	0.009**	
		(0.002)	(0.004)	(0.003)	(0.003)	(0.004)	
Credits		0.028***	0.030***	0.026***	0.031***	0.024***	
		(0.003)	(0.005)	(0.004)	(0.004)	(0.006)	
Technical Schools		-0.043	-0.020	-0.060	0.023	-0.159	
		(0.058)	(0.092)	(0.075)	(0.074)	(0.099)	
Lyceum		0.055	0.089	0.033	0.094	-0.027	
		(0.058)	(0.092)	(0.075)	(0.074)	(0.097)	
Late Enrollment		0.042	0.040	0.037	0.051	0.037	
		(0.052)	(0.091)	(0.062)	(0.066)	(0.084)	
Resident near University		0.008	-0.000	0.009	-0.050	0.099	
		(0.037)	(0.062)	(0.046)	(0.046)	(0.066)	
Observations	344	344	151	193	211	133	
Pseudo R-squared	0.017	0.069	0.068	0.072	0.091	0.048	
Log-likelihood	-1015	-960.8	-410.8	-548.5	-578.8	-376.7	

Notes: The Table reports marginal effects from Tobit estimates. Robust standard errors are reported in parentheses. The symbols \*\*\*, \*\*, \* indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

### 3.3. Heterogeneous Effects Across Students with Different Abilities

In this section we investigate whether the effect of treatment differs according to students' abilities. To divide students into a high ability and a low ability group, we have used three different measures of their skills: 1) the grades obtained at high school (*High School Grade*); 2) the number of credits gained by the beginning of the experiment (*Credits*); 3) their predicted performance (*CompositeAbility*).

To obtain the latter composite measure of student ability, we follow Angrist and Lavy (2009) and we first estimate a model for student performance considering exclusively students in the control group, using as explanatory variables individual characteristics such as gender, type of high school attended, high school grade. The estimated coefficients of this model are then used for predicting the performance of students in treated and control groups, on the basis of their effective characteristics.

In Table 5 are reported a number of probit estimates showing the effects of treatment on the probability of passing exams respectively on high and the low ability students. In column 1 we measure student ability considering *Credits* and to investigate whether there are heterogeneous effects we use the interaction variable *Credits\*Treatment*, where *Credits* is demeaned. The coefficient on the interaction term is positive and statistically significant at the 10 percent level, suggesting that the positive impact of more frequent exams increases with student ability.

**Table 5. Probit Estimates. Heterogeneous effects according to student abilities**

	(1) All	(2) High ab.	(3) Low ab.	(4) High ab.	(5) Low ab.	(6) High ab.	(7) Low ab.
Treatment (Intermediate Exam)	0.184*** (0.057)	0.343*** (0.083)	0.050 (0.053)	0.360*** (0.078)	0.058 (0.055)	0.215*** (0.081)	0.124** (0.053)
Treatment*Credits(demeaned)	0.021* (0.011)						
Macroeconomics	-0.196*** (0.057)	-0.168** (0.084)	-0.147*** (0.050)	-0.314*** (0.087)	-0.085 (0.056)	-0.177* (0.099)	-0.067 (0.071)
Female	-0.054 (0.058)	0.079 (0.092)	-0.142** (0.058)	0.048 (0.094)	-0.087 (0.060)	0.034 (0.086)	-0.108** (0.052)
High School Grade	0.013*** (0.004)	0.019*** (0.006)	0.005 (0.004)	0.011 (0.012)	0.003 (0.006)	0.015** (0.006)	0.004 (0.003)
Credits	0.023*** (0.007)	0.043*** (0.013)	0.012 (0.008)	0.042*** (0.009)	0.024*** (0.005)	0.038*** (0.009)	0.010 (0.006)
Technical Schools	0.024 (0.093)	-0.007 (0.147)	0.058 (0.092)	-0.204 (0.134)	0.191 (0.142)	0.036 (0.157)	0.029 (0.068)
Lyceum	0.131 (0.096)	0.164 (0.151)	0.030 (0.093)	-0.139 (0.136)	0.274** (0.122)	0.190 (0.160)	-0.048 (0.066)
Late Enrollment	0.004 (0.085)	-0.029 (0.131)	0.010 (0.085)	-0.257** (0.110)	0.197* (0.107)	-0.080 (0.123)	0.071 (0.094)
Resident near University	0.050 (0.056)	0.016 (0.095)	0.035 (0.053)	0.037 (0.092)	0.081 (0.056)	0.046 (0.087)	0.025 (0.056)
Observations	344	166	178	169	175	172	172
Pseudo R-squared	0.255	0.226	0.154	0.267	0.223	0.148	0.172
Log-likelihood	-165.710	-88.771	-70.899	-85.701	-72.155	-101.055	-61.918

Notes: The Table reports marginal effects. Robust standard errors are reported in parentheses. The symbols \*\*\*, \*\*, \* indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

To better describe this aspect we run separate regressions for students with ability above and

below the median according to the three measures of ability we have available. In columns 2 and 3 students are grouped according to the number of *Credits*, in columns 4 and 5 using the *High School Grade* and in columns 6 and 7 we define the High and Low ability groups on the basis of their predicted performance, *CompositeAbility*. Changing the measure of student ability does not change the results: the effect of treatment on high ability students is always stronger, while the effect for low ability students is not significantly different from zero.

The same findings emerge also when we use as dependent variable the grades obtained at exams (estimates are not reported and available upon request).

#### **4. Treatment effects on students effectively undertaking the intermediate exam: IV estimates**

In the previous section we have analyzed “intention-to-treat” effects, since we have considered all students in the treatment group independently on their effective participation at the intermediate exam. Therefore, the estimated effects are diluted by the fact that some treated students may actually not have undertaken the intermediate exam (see Angrist and Pischke, 2009; Bloom, 1984).

As a matter of fact, students in the treatment group may refuse treatment choosing to not sit for the intermediate exam or choosing after the intermediate exam to retake the entire exam. On the other hand, the 172 students assigned to the control group could not shift to the treatment group.

In this section we analyze the impact of the effective treatment defining two alternative treated groups. Firstly, we consider as treated the 133 students (out of 172) assigned to the treatment group who have effectively undertaken the intermediate exam regardless of whether they have subsequently shifted to the entire exam (“Effective Treatment 1”). Secondly, we define a more restricted treated group excluding from the 133 students undertaking the intermediate exam 26 students who have decided to take in July the entire exam (“Effective Treatment 2”).

In both cases, to deal with the endogeneity problems related to the effective treatment, we adopt an instrumental variable strategy using as an instrument for the effective participation the randomly assigned treatment status (*Treatment*).

In Table 6 are reported Instrumental Variable estimates. In columns 1 and 2 we consider as dependent variable *Pass*, reporting the marginal effects of an IV-probit model. In column 1 we focus on “Effective Treatment 1” while in column 2 we deal with “Effective Treatment 2”. In columns 3 and 4 the dependent variable is *Grade*. We report IV-tobit estimates (the marginal effects conditional on being uncensored) considering respectively “Effective Treatment 1” and “Effective Treatment 2”.

In Panel (B) are reported first stage results where we show that the assigned treatment status is highly significant in predicting the effective treatment.

Results show that adjusting “intention-to-treat” effects for non-participation leads to a stronger impact of the treatment on student performance measured both as the probability of passing the exam

and by the grade obtained. The increase in the effect is proportional to the reciprocal of the participation rate.

**Table 6. IV estimates of the effects of effective participation to the treatment**

	(1) Probit Pass	(2) Probit Pass	(3) Tobit Grade	(8) Tobit Grade
Effective Treatment 1	0.264*** (0.079)		6.446*** (0.991)	
Effective Treatment 2		0.341*** (0.098)		8.703*** (1.364)
Macroeconomics	-0.181*** (0.058)	-0.172*** (0.059)	-2.334*** (0.687)	-2.181*** (0.705)
Female	-0.053 (0.059)	-0.066 (0.060)	-0.602 (0.702)	-0.613 (0.714)
High School Grade	0.013*** (0.004)	0.012*** (0.004)	0.088** (0.043)	0.070* (0.043)
Credits	0.035*** (0.005)	0.034*** (0.005)	0.554*** (0.056)	0.529*** (0.056)
Technical Schools	-0.017 (0.091)	-0.031 (0.090)	-0.944 (1.042)	-1.346 (1.054)
Lyceum	0.106 (0.092)	0.095 (0.092)	1.121 (1.049)	0.819 (1.068)
Late Enrollment	0.023 (0.088)	0.033 (0.088)	0.818 (1.014)	1.249 (1.060)
Resident near University	0.047 (0.056)	0.048 (0.056)	-0.070 (0.694)	-0.153 (0.703)
Observations	344	344	344	344
Panel B: First Stage				
Treatment (Randomly Assigned)	0.773 (0.0317)	0.626 (0.036)	0.773 (0.0317)	0.626 (0.036)
First-Stage F-statistics (p-value)	594.92 (0.000)	294.93 (0.000)	594.92 (0.000)	294.93 (0.000)

Note. In the First Stage the dummy for the assigned treatment is used as an instrument for “Effective Treatment 1” and “Effective Treatment 2”. Standard errors (reported in parentheses) are corrected for heteroskedasticity. The symbols \*\*\*, \*\*, \* indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

## 5. Disentangling Feedback and Workload Division Effects

The positive impact of treatment on student performance emerging in previous estimates can be interpreted as the joint result of a feedback provision effect and the effects deriving from the workload division and commitment obtained defining more frequent deadlines.

In this section we propose a framework to try to disentangle the feedback provision effect from the other two effects. At this aim, instead of considering student performance on the whole examination, we have organized our data at student-module level to have information on the results obtained by students separately for the two modules of each examination (Microeconomics-1 and Microeconomics-2; Macroeconomics-1 and Macroeconomics-2).

We exploit the fact that treated students undertaking module 1 compared to control students undertaking the same module benefit exclusively from a workload division or commitment effect,

while no feedback effect is at work. Then, comparing the performance of treated and control students at this module we are able to disentangle the workload/commitment effect. Instead, to isolate the feedback effect, we compare the difference in performance of treated students between the two modules. Treated students in both modules benefit of workload/commitment effects, but only for the second module they can obtain positive effects from feedback. Then, the difference in the performance obtained by these students at the two modules should reflect feedback provision, given that we are able to neutralize any eventual heterogeneity in the difficulty level of the first and second module by subtracting the difference in performance at the two modules obtained by control students. Unfortunately, we are not able to distinguish the workload division effect from the commitment effect deriving from frequent deadlines since they overlap in each module.

Formally, we assume that student performance is determined as follows:

$$[3] \quad Y_{ij} = \alpha_0 + \alpha_F F_{ij} + \alpha_W W_i + \alpha_D D_2 + \phi X_i + \varepsilon_{ij}$$

where  $Y_{ij}$  is the performance of student  $i$  at module  $j$ , with  $j=1,2$ . The performance is affected by feedback provision,  $F_{ij}$ , which is a dummy equal to 1 for students receiving feedback (that is, treated students undertaking the second module), workload division or commitment,  $W_i$ , which is a dummy taking value of one for students undertaking the exam in two separated modules (treated students), a dummy for the second module,  $D_2$ , measuring the relative difficulty of this module, and a vector  $X_i$  of individual characteristics and a dummy for Macroeconomics. We expect that the workload/commitment has a positive effect on student performance,  $\alpha_W > 0$ , while according to the literature the feedback effect  $\alpha_F$  could be either positive or negative.

In the Table below using eq. [3] we report the expected performance of treated and control students respectively at module 1 and 2.

**Table 7. Student's expected performance by treatment status and module**

Treated, I module:	$E(Y_{T1}) = \alpha_0 + \alpha_W + \phi \bar{X}_i$	since $F_{ij} = 0, W_i = 1; D_2 = 0$
Treated, II module:	$E(Y_{T2}) = \alpha_0 + \alpha_F + \alpha_W + \alpha_D + \phi \bar{X}_i$	since $F_{ij} = 1, W_i = 1; D_2 = 1$
Control, I module:	$E(Y_{C1}) = \alpha_0 + \phi \bar{X}_i$	since $F_{ij} = 0, W_i = 0; D_2 = 0$
Control, II module:	$E(Y_{C2}) = \alpha_0 + \alpha_D + \phi \bar{X}_i$	since $F_{ij} = 0, W_i = 0; D_2 = 1$

For example, the expected performance for treated students at module 2 is equal to  $E(Y_{T2}) = \alpha_0 + \alpha_F + \alpha_W + \alpha_D + \phi \bar{X}_i$  as these students have received feedback, have a reduced workload and are undertaking the second module.



To disentangle the feedback and the workload/commitment effect we consider the following equations:

$$\begin{aligned} [E(Y_{T2}) - E(Y_{T1})] - [E(Y_{C2}) - E(Y_{C1})] &= \alpha_F && \text{which gives the "feedback effect"} \\ E(Y_{T1}) - E(Y_{C1}) &= \alpha_W && \text{which gives the "workload division/commitment effect"} \end{aligned}$$

where we exploit the fact that the characteristics  $(X_i)$  of treated and control students are on average the same thanks to the random assignment to the two groups.

We base our analysis on a sub-sample of 166 students, excluding from the sample: students who did not sit for the examinations (108 students: 22 treated and 86 control); non complier students, that is, treated students who have shifted to the whole examination (43); students who have taken the intermediate examination but have not taken the second part (27). We exclude these students since for them it is not possible to clearly disentangle the feedback and the workload effects following the framework presented above.

We investigate the effects produced by feedback and workload/commitment both on the grades obtained by students and on their probability of passing the examinations. Since we are using two observations for each student, standard errors have been clustered at student level.

For the sample of students considered in this analysis we do not have censored observations and, as a consequence, we use an OLS model to analyze the effects of interest on *Grade*. Estimates are reported in columns (1) and (2) of Table 8: in the first we do not control for individuals characteristics, which are instead included in the second specification. It emerges that the "workload/commitment" effects are positive and highly statistically significant. Students in the treated group obtain a grade higher of about 4 points (about 0.5 standard deviations of *Grade*). This effect can be interpreted both in relation to the splitting of class workload in two parts – which may help students at better organizing their studying activities – and in relation to the "no procrastination commitment" obtained thanks to the fact that students in the treatment group face more frequent deadlines that may induce them to not procrastinate effort.

On the other hand, the estimates show that the feedback effect is far from being statistically significant.

Similar results are obtained also when we consider as dependent variable the dummy *Pass* using a probit model. As shown in columns (3) and (4), respectively with and without individual controls, workload/commitment increases the probability of passing the exam of 23.1 percentage points (significant at the 1 percent level) while the feedback effect is null.

The control variables show signs and coefficients similar to those discussed in the previous sections.

The feedback and workload division effects estimated in this section can be related to the whole treatment effect estimated in the previous sections using as a measure of student performance

the average grade of module 1 and 2 ( $\bar{Y}_i = (Y_{i1} + Y_{i2})/2$ ). In fact, using equation [3] it is straightforward to show that the whole treatment effect is equal to:

$$E(\bar{Y}_T) - E(\bar{Y}_C) = \frac{(\alpha_0 + \alpha_L) + (\alpha_0 + \alpha_F + \alpha_W + \alpha_D)}{2} - \frac{(\alpha_0) + (\alpha_0 + \alpha_D)}{2} = \alpha_W + \frac{\alpha_F}{2}$$

Estimating by OLS the whole treatment effect (1 observation for each student) on the sub-sample of students considered to disentangle the two effects of interest, we obtain a coefficient of 4.14 ( $t$ -stat= 4.43), which corresponds to the coefficient on workload division (4.367) plus the coefficient (divided by 2) on feedback (-0.452/2).

**Table 8. Feedback and Workload Division Effects**

	(1) OLS Grade	(2) OLS Grade	(3) Probit Pass	(4) Probit Pass
Feedback	-0.452 (0.900)	-0.452 (0.910)	-0.003 (0.074)	0.006 (0.085)
Workload Division	4.058*** (1.152)	4.367*** (0.957)	0.196*** (0.074)	0.231*** (0.078)
Macroeconomics	0.638 (1.118)	-1.928* (1.124)	0.058 (0.069)	-0.069 (0.086)
Module 2	-1.523** (0.647)	-1.523** (0.654)	-0.011 (0.054)	-0.016 (0.061)
Female		-1.417 (0.990)		-0.110 (0.069)
High School Grade		0.195*** (0.060)		0.011** (0.004)
Credits		0.560*** (0.095)		0.027*** (0.007)
Technical Schools		1.205 (1.708)		0.011 (0.115)
Lyceum		2.690* (1.606)		0.062 (0.111)
Late Enrollment		-0.229 (1.525)		-0.012 (0.104)
Resident near University		-0.461 (1.066)		0.011 (0.073)
Constant	16.588*** (1.023)	-9.981* (5.551)		
Observations	332	332	332	332
R-squared	0.065	0.364		
Pseudo R-squared			0.031	0.169
Log-likelihood			-215.447	-184.648

Notes: Observations at module-student level: 2 observations for each student. In columns 3 and 4 we report marginal effects. Robust standard errors, clustered at student level, are reported in parentheses. The symbols \*\*\*, \*\*, \* indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

## 6. Are students substituting effort?

In this Section we investigate if treated students – thanks to the opportunity to undertake the intermediate examination – have provided more effort and achieved a better performance or whether they have focused on the targeted examinations devoting less effort in studying activities related to

other examinations.

In order to evaluate this aspect we analyze student performance on all the examinations students have to pass in the academic year considered. We stack data with the aim to use student-class level observations and deal separately with students of the first year (those attending Microeconomics) and students of the second year (students attending Macroeconomics).

In the first year students have to pass 8 examinations:<sup>5</sup> we have for these students 1544 observations (193 students \* 8 classes); in their second year students have to pass 8 examinations:<sup>6</sup> we end up with 1208 observations (151 students \* 8 classes). Our dependent variable, *Pass*, is a dummy equal to 1 when the student passed a given exam, and 0 otherwise.

We use a Linear Probability Model to estimate the following equation:

[4]

$$Pass_i = \beta_0 + \beta_1(Treatment_i) + \phi X_i + \psi_1 E_{1i} + \dots + \psi_7 E_{7i} + \pi_1(Treatment_i * E_{1i}) + \dots + \pi_7(Treatment_i * E_{7i}) + \varepsilon_i$$

where the probability of passing each examination is related to individual characteristics  $X$ , a dummy  $E_j$  ( $j=1..7$ ) for each examination (leaving as omitted category the eighth examination – Microeconomics or Macroeconomics according to the cohort of students considered) to control for unobserved factors such as the difficulty of the subject or the instructor's grading standard. The dummy *Treatment* allows us to identify the effect of treatment on the target examination ( $\beta_1$ ) while the interaction terms ( $Treatment * E_j$ ) inform us on whether treated students have a worse performance compared to control students in non targeted examinations. In fact, the performance of treated students in examination  $j$  is given by  $\beta_1 + \pi_j$ .

Estimates are reported in Table 9. In column 1 we report results considering first year students attending the Microeconomics class, while in column 2 are presented estimates for second year students attending the Macroeconomic class. The coefficients on *Treatment* are, as expected, in line with those shown in Table 3 (columns 3 and 4). On the other hand, we do not observe any significant difference among students in treated and control groups in the performance on non-target examinations. As shown at the foot of the Table, where are reported  $\beta_1 + \pi_j$  and the respective standard errors, the coefficients are never statistically significant.

Therefore, it seems that the improvement in student performance at the targeted examination is to be related to a higher student effort and is not driven by substitution effects.

---

<sup>5</sup> Business Administration 1, Public Law, French 1, Computer Sciences, English 1, Mathematics, Statistics, Microeconomics.

<sup>6</sup> Trade Law, Business Administration 2, Private Law, French 2, English 2, Financial Mathematics, Accounting, Macroeconomics.

**Table 9. The probability of Passing Targeted and Non-Targeted Exams. Linear Probability Model**

	(1)	(2)
	First Year	Second Year
Treatment	0.131* (0.070)	0.260*** (0.069)
Treatment*Exam 1	-0.110 (0.074)	-0.248*** (0.075)
Treatment*Exam 2	-0.154 (0.095)	-0.362*** (0.091)
Treatment*Exam 3	-0.131* (0.078)	-0.236*** (0.077)
Treatment*Exam 4	-0.090 (0.078)	-0.256** (0.099)
Treatment*Exam 5	-0.161** (0.082)	-0.258*** (0.092)
Treatment*Exam 6	-0.136* (0.076)	-0.367*** (0.088)
Treatment*Exam 7	-0.117 (0.080)	-0.239*** (0.080)
Exam 1	0.433*** (0.051)	-0.147*** (0.046)
Exam 2	-0.082 (0.067)	0.507*** (0.061)
Exam 3	0.443*** (0.057)	-0.053 (0.046)
Exam 4	0.412*** (0.058)	0.493*** (0.072)
Exam 5	0.247*** (0.061)	0.627*** (0.060)
Exam 6	-0.155*** (0.056)	0.427*** (0.069)
Exam 7	-0.227*** (0.056)	0.133** (0.055)
Implied Treatment Exam 1	0.021 (0.040)	0.012 (0.031)
Implied Treatment Exam 2	-0.023 (0.070)	-0.101 (0.078)
Implied Treatment Exam 3	0.000 (0.041)	0.024 (0.052)
Implied Treatment Exam 4	0.041 (0.041)	0.004 (0.074)
Implied Treatment Exam 5	-0.030 (0.063)	0.002 (0.064)
Implied Treatment Exam 6	-0.005 (0.065)	-0.107 (0.078)
Implied Treatment Exam 7	0.014 (0.061)	0.022 (0.074)
Observations	1544	1208
R-squared	0.344	0.326

Notes: The dependent variable is the dummy *Pass*. Standard errors (reported in parentheses) are corrected for heteroskedasticity and clustered at the student level. In all the regressions we control for Female, High School Grade, Type of High School attended, Late enrollment, Residence near the university. The symbols \*\*\*, \*\*, \* indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

## **7. Concluding Remarks**

Policymakers and researches often debate about students' performance and on policies that may help at boosting them. The improvement of students' performance is important both when considering primary and secondary education and when looking at tertiary education since skills acquired by undergraduate students are crucial for their success in the labor market. In addition, since in many countries, and particularly in Italy, tertiary education is characterized by high drop out rates and excessive duration of the academic career, the improvement of students performance may translate in a reduction of the number of students dropping out from university studies.

A widely investigated strategy to improve student performance is that based on educational resources, but a considerable impact may derive also from changes in the organization of teaching activities and in evaluation practices. Recently, a number of papers have shown that providing students with feedback information both on their interim performance than on the performance of their peers helps at increasing student results (Azmat and Iriberry, 2010, Bandiera et al. 2009). In this vein, the aim of our paper has been to investigate the effect of more frequent examinations allowing students to obtain a number of beneficial effects, deriving both from the provision of feedback and from the division of the class workload and the commitment allowed by recurrent deadlines.

To analyze these effects we have carried out a randomized experiment involving undergraduate students enrolled at an Italian University and attending two introductory economics classes. Students included in the treated group were allowed to undertake an intermediate exam and were informed about the results obtained, while students in the control group were allowed to undertake exclusively the final examination at the end of the classes (as established by the University rules).

From our analysis it emerges that students undertaking the intermediate examination obtain a better performance both in terms of probability of passing the exam and of grades obtained. Treated students have a probability of 20 percentage points higher of passing the exam and their grades are 4-5 points higher. These positive effects are mainly concentrated among students endowed with higher abilities. The better performance obtained by treated students at the targeted examination seems not to be driven by substitution effects.

The design of our experiment allowed us to disentangle the effect deriving from more frequent examinations in a "feedback provision" effect and a "workload division or commitment" effect. Our estimates show that the positive impact of the intermediate examination is entirely due to the workload division or commitment effect, while it turns out that the feedback provision has no positive effect on performance.

According to our results the recent law that in Italy has reorganized classes in longer modules defining a maximum number of examinations may lead to an increase in the duration of students academic career.

## References

- Aoyagi M. (2010), Information feedback in a dynamic tournament, *Games and Economic Behaviour*, forthcoming.
- Azmat G. and Iriberri, N. (2010), The importance of relative performance feedback information: Evidence from a natural experiment using high school students, *Journal of Public Economics*, 94(7-8), pp. 435-452
- Bandiera, O., Larcinese, V., Rasul, I., (2008), “Blissful Ignorance? Evidence from a Natural Experiment on the Effect of Individual Feedback on Performance”, Mimeo.
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1986). Effects of frequent classroom testing (Report No. TM-860-524). San Francisco, CA: Paper presented at the Annual Meeting of the American Educational Research Association. (ERIC Document Reproduction Service No. ED 274 672).
- Beaulieu, R., Frost, B. (1989), “Impact of examination frequency on Achievement and Effect”, *Journal of Instructional Psychology*, 16, 145-150.
- Ederer, F. (2010), Feedback and motivation in dynamic tournaments, *Journal of Economics & Management Strategy*.
- Eriksson, T., Poulsen, A., Villeval, M. (2009), Feedback and incentives: experimental evidence. *Labour Economics*, 16, pp. 679–688.
- Freeman, R. and Gelber, A. (2010), Prize Structure and Information in Tournaments: Experimental Evidence, *American Economic Journal: Applied Economics*, 2(1), pp. 149-164 (16).
- Gibbs, M. (1991), An Economic Approach to Process in Pay and Performance Appraisals, Harvard Business School; Mimeo.
- Goltsman, M., Mukherjee, A. (2006), Information disclosure in multistage tournaments. Working paper, University of Western Ontario and Bates White LLC.
- Gürtler O. and C. Harbring, (2010), Feedback in Tournaments Under Commitment Problems: Experimental Evidence, *Journal of Economics & Management Strategy*, 19(3), pp. 771-810.
- Hattie J., H. Timperley (2007), The power of feedback, *Review of Educational Research*, 77(1), pp. 81-112.
- Kika, F. M., McLaughlin, T. F., & Dixon, J. (1992). Effects of frequent testing of secondary algebra students. *Journal of Educational Research*, 85(3), 159-162.
- Perry, M., Reny, P.J. (1999), On the failure of the linkage principle in multi-unit auctions, *Econometrica*, 67, 895–900.
- Tuckman, B., (1997), Using tests as an incentive to motivate procrastinators to study”, *Journal of Experimental Education*, 66, 141-147.
- Wang, R., Zhang, J. (2007), The role of information revelation in elimination contests, Working paper, Queens University.
- Yildirim, H. (2005), Contests with multiple rounds, *Games and Economic Behavior*, 51 (1), pp. 213–227.